



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Eye-guidance during real-world scene search

Citation for published version:

Nuthmann, A & Malcolm, GL 2016, 'Eye-guidance during real-world scene search: The role color plays in central and peripheral vision' *Journal of Vision*, vol. 16, no. 2, 3, pp. 1-16. DOI: 10.1167/16.2.3

Digital Object Identifier (DOI):

[10.1167/16.2.3](https://doi.org/10.1167/16.2.3)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Journal of Vision

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Eye guidance during real-world scene search: The role color plays in central and peripheral vision

Antje Nuthmann

Psychology Department, School of Philosophy,
Psychology and Language Sciences, University of Edinburgh,
Edinburgh, United Kingdom



George L. Malcolm

Department of Psychology, The George Washington University,
Washington, DC, USA



The visual system utilizes environmental features to direct gaze efficiently when locating objects. While previous research has isolated various features' contributions to gaze guidance, these studies generally used sparse displays and did not investigate how features facilitated search as a function of their location on the visual field. The current study investigated how features across the visual field—particularly color—facilitate gaze guidance during real-world search. A gaze-contingent window followed participants' eye movements, restricting color information to specified regions. Scene images were presented in full color, with color in the periphery and gray in central vision or gray in the periphery and color in central vision, or in grayscale. Color conditions were crossed with a search cue manipulation, with the target cued either with a word label or an exact picture. Search times increased as color information in the scene decreased. A gaze-data based decomposition of search time revealed color-mediated effects on specific subprocesses of search. Color in peripheral vision facilitated target localization, whereas color in central vision facilitated target verification. Picture cues facilitated search, with the effects of cue specificity and scene color combining additively. When available, the visual system utilizes the environment's color information to facilitate different real-world visual search behaviors based on the location within the visual field.

Introduction

We live in a colorful world in which we regularly need to locate specific objects—for example, keys on a desk or a car in the parking lot. Given the rich complexity of our environment, how do we locate necessary items quickly? The human visual system is a

limited resource with high resolution at the fovea and diminishing acuity toward the periphery (Strasburger, Rentschler, & Jüttner, 2011; Wilson, Levi, Maffei, Rovamo, & DeValois, 1990). The resolution of the visual system changes gradually and systematically from the central fovea into the periphery rather than suddenly, though researchers commonly divide the visual field into three major regions: foveal, parafoveal, and peripheral. The foveal region extends out to an angle of eccentricity of 1°, the parafoveal region from 1° to 4° to 5°, and the peripheral region encompasses the remainder of the visual field. The fovea and parafovea together are commonly referred to as *central vision*. Critically, these different visual field regions serve different functions in visual search (Cornelissen, Bruin, & Kooijman, 2005; Nuthmann, 2014).

A question that arises from the functional segregation of the visual field is what properties are extracted from the periphery and from fixation to facilitate the search process. The present study investigates how color information, in particular, is extracted and utilized from different regions of the visual field. Color information has been previously shown to play a critical role in search, dominating other features—such as orientation and size (Hannus, van den Berg, Bekkering, Roerdink, & Cornelissen, 2006; Rutishauser & Koch, 2007; Williams, 1967)—facilitating guidance (Wolfe, 1994) and improving computational models of attention (Itti & Koch, 2000) and real-world search (Hwang, Higgins, & Pomplun, 2009; Zelinsky, 2008).

However, despite this well-documented importance, the effect of color across different regions of the visual field is less well understood. Color sensitivity is not consistent across the visual field; it is best in the fovea and then declines in the periphery (Hansen, Pracejus, & Gegenfurtner, 2009; Mullen, 1991; Mullen & Kingdom, 1996). Color perception persists in the intermediate

Citation: Nuthmann, A., & Malcolm, G. L. (2016). Eye guidance during real-world scene search: The role color plays in central and peripheral vision. *Journal of Vision*, 16(2):3, 1–16, doi:10.1167/16.2.3.

doi: 10.1167/16.2.3

Received June 1, 2015; published January 29, 2016

ISSN 1534-7362

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.



Downloaded From: <http://jov.arvojournals.org/pdfaccess.ashx?url=/data/Journals/JOV/934904/> on 04/29/2016

periphery; for example, chromatic detection is still possible at an eccentricity of 50° of visual angle (Hansen et al., 2009). The question, therefore, is how color information at fixation and in the periphery affects the search process. Color, for instance, could facilitate scanning for a target by improving saccadic guidance to items in the visual periphery, or could facilitate target verification by facilitating the recognition process, or both to varying degrees.

Early visual search research focused on identifying features that benefitted search more than revealing how these properties were utilized by the visual system. Studies used simple displays, demonstrating that search for a target defined by its color can be highly efficient. Color singletons “pop out” in search displays, and detection time remains fast regardless of the number of distractors present (D’Zmura, 1991; Nagy & Sanchez, 1990; Treisman & Gelade, 1980). Color singletons even capture attention irrespective of the observer’s attentional set (Theeuwes, 1994). Color also provides top-down guidance to locate items in conjunctive search tasks (Wolfe, 1994). Subsequent studies that tracked eye movements in simple displays found that knowing the color of the target biased saccades toward items that had the same color (Hannus et al., 2006; Rutishauser & Koch, 2007; Williams, 1967).

While research that tracked participants’ eye movements during search through simple displays began to demonstrate that peripheral color information facilitates attentional guidance, these stimuli lacked the rich complexity of real-world scenes. Scenes simultaneously contain more noise (e.g., clutter; Rosenholtz, Li, & Nakano, 2007) and more information (e.g., scene context; Malcolm & Henderson, 2010; Neider & Zelinsky, 2006) than object arrays. However, only a few recent studies have begun to explore the role of color in real-world scene perception and search. These studies have approached color’s contribution in three different ways.

The first has been to present naturalistic scene images in either color or grayscale in order to assess how removing color information affects task performance and global eye-movement parameters. An exploratory study by Hwang, Higgins, and Pomplun (2007) applied this approach to scene search. In order to minimize semantic effects during search, the authors used small scene cutouts as targets and rotated both the scene and the previewed target by 90°, 180°, or 270°. Under these conditions, search proved to be difficult, with more accurate and faster target detection in the color than in the grayscale condition. Interestingly, mean fixation duration was not affected by the color manipulation. However, in other studies using similar color and grayscale manipulations, mean fixation durations were found to be longer for grayscale than for color scenes in both a scene memorization task (von

Wartburg et al., 2005) and a free viewing task (Ho-Phuoc, Guyader, Landragin, & Guerin-Dugue, 2012). All three studies consistently reported that saccade amplitudes did not differ for color and grayscale scenes (Ho-Phuoc et al., 2012; Hwang et al., 2007; von Wartburg et al., 2005).

A second approach has been to present observers with full-color scene images and analyze systematic associations between local color information and points of fixation. Tatler, Baddeley, and Gilchrist (2005), using a scene memorization task, found that for chromaticity, discrimination of fixated and nonfixated scene regions was at 57% at its highest (at a spatial scale of 10.8 cpd). While this was significantly above chance (chance being 50%), edge content and contrast were stronger contributors to fixation selection than color. However, other research has found that local color properties (i.e., lightness and the red–green and blue–yellow chromatic components) have a stronger influence on fixation selection (Amano & Foster, 2014; see also Amano, Foster, Mould, & Oakley, 2012). In the study by Amano and colleagues, participants had 1 s to detect small (0.25°) targets disks that were embedded in natural scene images. Local color explained more variance in fixation positions than lightness, lightness contrast, and edge density (Amano & Foster, 2014).

The third approach involves implementing computational models of visual salience that attempt to predict fixation locations (for reviews, see Borji & Itti, 2013; Borji, Sihite, & Itti, 2013). In these models, visual differences are assumed to be informative and attract attention and gaze. An early saliency model (Itti & Koch, 2000; Itti, Koch, & Niebur, 1998) had the activation map computed from three basic features: luminance, color, and orientation. The work by Jost, Ouerhani, von Warburg, Müri, and Hügli (2005) explicitly investigated the role of chromatic features in the saliency model. Participants freely viewed color images of scenes, fractals, and abstract art images. Fixation patterns were compared with two different versions of a saliency model—a gray-level model and a color model—with the color model performing better of the two. The results thus suggested that color information has a noticeable influence on human visual attention.

These approaches are not always mutually exclusive. Frey, Honey, and König (2008) combined experimental and computational approaches, with observers freely viewing color or grayscale scene images with fixation placements differing across the two conditions. Additional analyses of three color features (saturation, red–green color contrast, and blue–yellow color contrast) found that effects of color on fixation selection in scenes were category specific: *Rainforest* was the only category for which all color features were salient (for a follow-up, see Frey et al., 2011). The saliency model

(Itti et al., 1998) did not reproduce the human behavior to the same extent.

Despite the range of studies that have examined the effect of color on real-world gaze control, there are still critical limitations to our knowledge of color's influence on search. For one, the empirical studies that used a real-world scene search task did not have participants look for target objects embedded in a scene but rather for cued scene patches (Hwang et al., 2007) or small disks (Amano & Foster, 2014; Amano et al., 2012). While these studies still invoked some of the top-down guidance factors associated with search, they may have artificially biased color information's contribution in the absence of other factors. Real-world search for objects relies not only on target features (Bravo & Farid, 2009; Malcolm & Henderson, 2009) but also on the target's semantic relationship with the scene (Eckstein, Drescher, & Shimozaki, 2006; Malcolm & Henderson, 2010; Neider & Zelinsky, 2006; Spotorno, Malcolm, & Tatler, 2014, 2015). Additionally, objects within a scene bias gaze allocation (Foulsham & Kingstone, 2013; Malcolm & Shomstein, 2015; Nuthmann & Henderson, 2010; Pajak & Nuthmann, 2013). The present study therefore used target objects embedded in scene images.

An additional limitation of previous research using scene statistics and computational models is that they rely on correlation techniques. In order to establish causality in the current study, we manipulated color information in central and peripheral vision. In the extreme cases, a search scene was presented in either full color or grayscale. In the full-color condition (C), color was available in both central and peripheral vision. In the full-grayscale condition (G), color was removed from both central and peripheral vision. We expected search to be more efficient in the full-color than in the full-grayscale condition (Hwang et al., 2007). However, any observed costs could be due to the removal of color in central vision, peripheral vision, or both. In order to disentangle these possibilities, we added two experimental conditions that selectively removed color information in either central vision (C-G) or peripheral vision (G-C) using the gaze-contingent "moving window" technique (Loschky & McConkie, 2002; McConkie & Rayner, 1975; Nuthmann, 2013). By removing color information from a specific region of the visual field and recording the ensuing search deficits, we can reverse correlate how the visual system utilizes color from that region. Removing color information from peripheral vision could, for instance, impede image segmentation and thereby saccade target selection (cf. Foulsham & Underwood, 2011). Therefore, we hypothesized that when color is not available in peripheral vision (G-C and G conditions), it should take longer to locate the target in the scene. On the other hand, the process of recognizing the target may

be prolonged when color is not available in central vision (C-G and G conditions). This would corroborate with surface-plus-edge-based hypotheses of object recognition, which suggest that both edges and color contribute to recognition (Humphrey, Goodale, Jakobson, & Servos, 1994; Wurm, Legge, Isenberg, & Luebker, 1993). In contrast, edge-based theories of object recognition minimize the role of color information (Biederman & Gerhardstein, 1993) and are supported by some empirical results (Biederman & Ju, 1988; Davidoff & Ostergaard, 1988; Ostergaard & Davidoff, 1985). In this case, there should be minimal difference in target verification times between the C-G and G conditions and the C and G-C conditions.

Finally, we manipulated how much top-down information viewers had by providing either a word or a picture cue of the target prior to search. When a picture cue is presented, the visual features are remembered and contrasted against features in the search image, with regions of high correlation attracting attention (Desimone & Duncan, 1995). The more features that are known, the better the resulting search. Therefore, an exact picture of the target would provide the most benefit because it contains all of the target's exact visual features, and a word cue would provide the least benefit because it contains none of the exact visual features. This process of matching features stored in visual working memory to those in an image to direct search is known as target template guidance and has been demonstrated to facilitate gaze behavior during real-world search (Malcolm & Henderson, 2009). As a final analysis, we sought to tease out the respective influence that target template color had on search by taking advantage of our parametric manipulation of scene color. In the current experiment, picture cues contained all of the target's features, including color. Thus, when viewing full-color images, all of the target template's features could be contrasted with those in a scene to locate the target, providing the maximum facilitation for search. However, when a picture cue is followed by a full-grayscale image, all of the template's features can be contrasted with those in the scene to direct search except for color, which will be mismatched (e.g., if the cue is a picture of a green bowl and the scene is in grayscale, observers can still use the cue's size, shape, and so on to facilitate search, but knowing that the object is green will be useless). We can take advantage of this mismatch to identify which search behaviors rely heavily on color information by observing what processes are significantly inhibited when a picture is presented prior to a full-grayscale scene.

As discussed above, we expected that in general full-color scenes should facilitate faster searches than full-grayscale scenes regardless of the cue. Likewise, we expected that picture cues will result in faster searches

than word cues (Malcolm & Henderson, 2009, 2010; Spotorno et al., 2014). However, if a particular search component relies heavily on a target template's color information, then we should find an interaction with much greater facilitation for picture cues over word cues in full-color scenes than in full-grayscale scenes. When a picture cue is presented prior to a full-grayscale scene, the mismatching color will cause the respective search process to be strongly inhibited. Conversely, if a template's color is one of many features used to direct real-world search, then we should find no interaction between scene color and cue type. When a picture cue is presented prior to a full-grayscale scene, the lack of color information will slow search, but viewers can match other features (e.g., luminance, shape) to the template to direct guidance.

Method

Participants

Thirty-two participants (15 males, 17 females) between the ages of 17 and 33 years (mean age = 21.3 years) participated in the experiment. All participants had normal color vision and normal or corrected-to-normal visual acuity by self-report. They were compensated at a rate of £6 per hour for their participation. Participants gave their written informed consent prior to the experiment, which conformed to the tenets of the Declaration of Helsinki.

Apparatus

Stimuli were presented on a 21-in. cathode ray tube monitor with a refresh rate of 140 Hz at a viewing distance of 90 cm, taking up a $24.8^\circ \times 18.6^\circ$ (width \times height) field of view. A chin rest was used to keep participants' head position stable. During stimulus presentation, participants' eye movements were recorded with an SR Research (Kanata, Ontario, Canada) EyeLink 1000 desktop mount system. It was equipped with the 2000-Hz camera upgrade, allowing for binocular recordings at a sampling rate of 1000 Hz for each eye. The system's average spatial accuracy ranges between 0.25° and 0.5° of visual angle. The experiment was implemented in MATLAB 2009b (The MathWorks, Natick, MA) using the OpenGL-based Psychophysics Toolbox 3 (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007), which incorporates the EyeLink Toolbox extensions (Cornelissen, Peters, & Palmer, 2002). The software allowed precise timing control over the display changes.

Stimulus materials

Stimuli consisted of 80 colored photographs of real-world scenes from a variety of categories (indoor and outdoor, natural and manmade). Each scene was scaled to 800×600 pixel resolution. One object (e.g., scissors) was chosen as the search target for each scene. Criteria for target selection included the following: The target was not completely monochrome, occurred only once in the scene, did not appear in the center of the image, and was not occluded. Search objects had an average size of $3.5^\circ \times 3.2^\circ$ (width \times height) and an average width:height ratio of 6:5. Because we used the eye-movement data to verify that target objects had indeed been found, there were no target-absent trials.

To create the picture cues, the target objects were copied and pasted into a blank background using Adobe Photoshop CS (Adobe, San Jose, CA). Picture cues were edited so that they did not contain any of the surrounding scene image and were then positioned at the center of an 800×600 pixel gray background image. A further 80 corresponding word cues were created that contained only the names of the target objects, presented in a 72-point font subtending 2.14° in height centered within the same gray background. Images were transformed to grayscale using the MATLAB function *rgb2gray*, which eliminates the hue and saturation information while retaining the luminance.

Design

In a 4×2 design, the availability of color across the visual field (color, color-gray, gray-color, gray) was crossed with search cue (word vs. picture). The availability of color was manipulated such that scenes were presented in four color conditions: (1) full color (C), (2) color in peripheral vision and gray in central vision (C-G), (3) gray in peripheral vision and color in central vision (G-C), and (4) grayscale (G). The presence of color in central or peripheral vision was manipulated using a 5° gaze-contingent moving window that followed participants' gaze. Both in the real world and in our laboratory setup, the low-resolution periphery is much larger than the high-resolution central area around fixation. In the experiment, the central area covered 17% of the image area; consequently, the peripheral area covered 83%. For statistical analysis, the four levels of the factor "scene color" were ordered according to the amount of color information available in the scene in descending order from full color to no color (C: 100%; C-G: 83%; G-C: 17%; G: 0%). Figure 1 provides a visualization of the color manipulation and ordering of conditions. The color conditions were crossed with a manipulation of

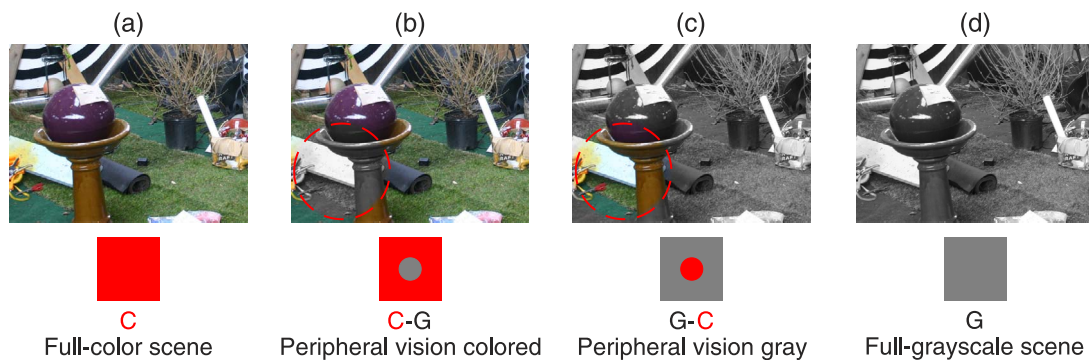


Figure 1. Each scene was presented in one of four color conditions: (a) full color (**C**), (b) color in peripheral vision and gray in central vision (**C-G**), (c) gray in peripheral vision and color in central vision (**G-C**), or (d) grayscale (**G**). The presence or absence of color in central and peripheral vision was manipulated using a 5° gaze-contingent window that followed participants' gaze. The red circles in panels b and c denote the window boundary; they are provided for clarity and were not present in the experiment. To summarize, color was either available in peripheral vision (**C** and **C-G** = panels a and b) or removed from peripheral vision (**G-C** and **G** = panels c and d). Likewise, color was either available in central vision (**C** and **G-C** = panels a and c) or removed from central vision (**C-G** and **G** = panels b and d).

the search cue; the search object was cued with either a word label or a picture of the target.

To control for item effects, we combined a set-matching procedure with counterbalancing of items across conditions. The 80 scene items were assigned to eight lists of 10 scenes each (see Supplementary Figure S1). For 22 scenes, search-time data from word cue/full-color conditions were available from previous studies (Nuthmann, 2013, 2014). These scenes were assigned to scene lists (two or three per list) such that mean search time was comparable for each list. Otherwise, scenes were algorithmically assigned such that the mean values for the following variables were comparable across scene lists: length of the word cue (characters), width, height, and area covered by the search object. The scene lists were rotated over participants such that a given participant was exposed to a list for only one of the eight experimental conditions created by the 4×2 design. There were eight groups of four participants, and each group of participants was exposed to unique combinations of list and experimental condition. To summarize, participants viewed each of the 80 scene items once, with 10 scenes in each of the eight experimental conditions. Across the 32 participants, each scene item appeared in each condition four times.

The search cue manipulation was blocked so that participants completed two blocks of trials in the experiment: one block of scenes with word search cues and one block with picture search cues. The order of word and picture search cues blocks was counterbalanced across subjects. Within each of those blocks, the scene color manipulation was implemented in four blocks (corresponding to the four color conditions), while those subblocks were randomized within a given

search cue block. Each subblock started with a practice trial.

Creation of gaze-contingent moving windows

For gaze-contingent presentation, full-color and full-grayscale images were merged in real time using alpha blending. In order to selectively manipulate color information in central and peripheral vision, we used a circular alpha mask with a 5° radius. The perimeter of the circular mask was smoothed with a Gaussian low-pass filter (Nuthmann, 2013). Specifically, the mask was a circular 1-center, 0-surround map, which was centered at gaze location. Pixel values of 1 (white) represent portions of the foreground image that show through, whereas values of 0 (black) are masked and therefore replaced by the corresponding background image pixels. Therefore, for the **C-G** condition (Figure 1b), the color scene was used as background image and the grayscale version of the scene as foreground image. Conversely, in the **G-C** condition, the color scene served as foreground image and the grayscale scene served as background image (Figure 1c). A detailed account of the gaze-contingent implementation is provided in Nuthmann (2013). Supplementary Movie S1 shows an exemplary trial from the word cue **G-C** condition.

Procedure

At the beginning of the experiment, a 9-point calibration procedure was performed, followed by a 9-point calibration accuracy test (validation). The trial structure was as follows. At the beginning of each trial

a fixation cross was presented at the center of the screen for 600 ms and acted as a fixation check. The fixation check was deemed successful if gaze position, averaged across both eyes, continuously stayed within an area of 40×40 pixels ($1.24^\circ \times 1.24^\circ$) for 200 ms. If this condition was not met, the fixation check timed out after 500 ms. In this case, the fixation check procedure was either repeated or replaced by another calibration procedure. If the fixation check was successful, the cross was replaced with the search cue, which was either a word identifying the target or an exactly matching picture of the target. The search cue remained on the screen for 1.5 s and was followed by a central fixation cross for 500 ms and then the scene. Once subjects located the target object, they should fixate their gaze on it and press a button on the controller to end the trial. Trials timed out 15 s after stimulus presentation if no response was made. There was an intertrial interval of 1.5 s before the next fixation cross was presented.

Data analysis

Saccades were defined with a $50^\circ/\text{s}$ velocity threshold using a nine-sample saccade detection model. Raw data were converted into a fixation sequence matrix using SR Research Data Viewer. Analyses of fixation durations and saccade lengths excluded fixations that co-occurred with blinks. Analysis of fixation durations disregarded fixations that were the first or last fixations in a trial. Fixations that had durations of less than 50 ms or greater than 750 ms were also excluded (0.87%), based on the assumption that they are not determined by online cognitive processes.

Linear mixed-effects models

In order to evaluate effects of scene color and search cue on search performance and eye-movement measures, the data were analyzed with either linear mixed-effects models (LMM; Baayen, Davidson, & Bates, 2008) or generalized linear mixed-effects models (GLMM; Jaeger, 2008). LMMs were used for continuous response variables (e.g., search time). Search success or accuracy, a binary variable, was analyzed with a binomial GLMM with a logit link function.

LMMs are regression techniques, and experimental design factors therefore enter the model as contrasts (Kliegl, Wei, Dambacher, Yan, & Zhou, 2011). For a given dependent variable, contrasts were chosen such that they tested hypotheses about the expected pattern of means and provided interpretable information. For contrast coding, we used the nomenclature and example code by the UCLA Statistical Consulting Group (2011). For factor cue specificity, simple coding

(reference: word cues) was used such that the difference between the two experimental conditions is expressed relative to the grand mean. To test effects of scene color on a given dependent variable, hypothesis-driven planned contrasts were specified. Moreover, incremental model building and model comparison was used to test whether scene color, search cue, and their interaction had significant effects. For each dependent variable, four LMMs were specified. The full model included color, cue, and their interaction as fixed effects. The color–cue model included color and cue as fixed effects. The color model included color as fixed effect only, and the cue model included cue as fixed effect only. Three model comparisons were conducted using likelihood ratio tests. The logic is as follows: If the full model provides a better fit than the cue model, then color has an effect (main effect or interaction); if the full model provides a better fit than the color model, then cue has an effect; and if the full model provides a better fit than the color–cue model, then there is a significant interaction of color and cue. To preview the results, the default pattern was that both color and cue had significant effects, with no significant interaction between the two. Therefore, if not otherwise stated, results from the color–cue model are reported in the Results and discussion section.

Mixed models are statistical models that incorporate both fixed-effects parameters and random effects (Baayen et al., 2008). Our LMMs included subjects (subject ID) and scenes (scene ID) as random effects to capture variance attributed to the randomness of subject and item sampling. LMMs minimize the false positives when they include the maximal random effects structure justified by the design (Barr, Levy, Scheepers, & Tily, 2013). For the full fixed-effects model, the maximal random effect structure would require estimating 72 parameters. For the color–cue model, the variance–covariance matrix of the random effects comprises 30 variance components and correlation parameters. For the subject factor, there were five variance components (intercept plus three within-subject color contrasts plus one within-subject cue contrast) and 10 correlation parameters for the possible correlations between each pair of these five components; the same 15 variance components and correlation parameters were included for the random factor “scene item.” Whenever a model did not converge, complexity of the random-effects structure was reduced by stepwise removal of random effects (Baayen et al., 2008; Barr et al., 2013).

Like all regression models, mixed models rely on the assumption that the model residuals are normally distributed, though this assumption is generally least important (Gelman & Hill, 2007). Diagnostics of the normality of residuals suggested that a log transformation would be appropriate for search times and

verification times but not for search initiation and scanning times. Therefore, for consistency, the four temporal measures of search efficiency were modeled on the original scale. Fixation durations were log transformed to achieve near-normal distribution of the residuals.

LMM analyses were conducted using the *lmer* program of the *lme4* package (Bates, Maechler, Bolker, & Walker, 2015) supplied in *R* (version 3.2.0; R Development Core Team, 2015). For model parameter estimation, maximum likelihood estimation was used. For the coded contrasts, coefficient estimates (b), their standard errors (SE), and t values ($t = b/SE$) are reported. An absolute t value of 1.96 or greater indicates significance at the alpha level of 0.05 (Baayen et al., 2008). For the GLMM, z scores and p values are provided.

Results and discussion

The results are presented in two main sections. The first section reports behavioral indices reflecting search efficiency, in particular search times. The second section provides an analysis of the observed eye-movement behavior. In particular, oculomotor behavior is used to decompose search times into behaviorally defined subprocesses of search (Castelhano, Pollatsek, & Cave, 2008; Malcolm & Henderson, 2009; Nuthmann, 2014; Spotorno et al., 2015). In addition, we examine saccade amplitude and fixation duration across the viewing period as general eye-movement measures.

Search performance

A trial was scored as accurate if the participant indicated to have located the target by button press and his or her gaze was within the target interest area. The target object was defined as the smallest possible rectangle to encompass the target object. For gaze scoring, 0.5° was added to either side of the rectangle to accommodate the spatial (in)accuracy of the eye tracker. To assess global search accuracy we report a binomial GLMM that included the intercept as fixed effect and by-subject and by-item random intercepts. Converting the parameter estimate for the intercept ($b = 3.2$, $SE = 0.18$, $z = 18.15$, $p < 0.001$) back from the log-odds scale to a probability reveals that the probability of correctly locating and identifying the target was very high at 0.96. Being at ceiling, search accuracy was modulated by neither scene color nor cue specificity.

Search time was defined as the elapsed time between search scene onset and the button press terminating the search. Only correct trials were included in the analysis. We expected search times to be faster for picture cues than for word cues (Malcolm & Henderson, 2009, 2010; Spotorno et al., 2014). Moreover, full-grayscale scenes should be associated with longer search times than full-color scenes (Hwang et al., 2007), with the C-G and G-C conditions falling somewhere in between. To test the effect of scene color on search times we used *backward difference coding* (also known as sliding differences or repeated contrasts). Here, the mean of the dependent variable for one level of the categorical variable is compared with the mean of the dependent variable for the prior adjacent level. To test the effect of cue specificity, we used *simple-coding*. The color-cue model best fit the data; it had color and cue as fixed effects as well as the maximal random effects structure. Given the contrast coding used, the intercept in the model reflects the grand mean estimated for the search-time data. The regression coefficient for cue specificity can be directly interpreted as the mean difference in search times between word and picture cue conditions. The three contrasts related to the scene color manipulation describe the differences in search times between (a) C-G and C, (b) G-C and C-G, and (c) G and G-C. In the LMM, the estimated grand mean was 1795 ms ($SE = 90$, $t = 20.02$). There was a significant speedup when picture cues were provided ($b = -260$, $SE = 62$, $t = -4.21$). With regard to the scene color manipulation, Figure 2a shows that search times numerically increase across the four color conditions. When color was removed from central vision (C-G), search times significantly increased by 281 ms ($SE = 70$, $t = 4.02$) compared with when the entire scene was presented in color (C). Removing color from the periphery (G-C) as opposed to central vision (C-G) was associated with a numerical increase in search time, but this failed to be significant ($b = 131$, $SE = 73$, $t = 1.79$).¹ No significant search-time costs occurred when removing color from the entire scene (G) as opposed to the periphery only (G-C) ($b = 112$, $SE = 77$, $t = 1.46$).

Eye-movement behavior

We used participants' gaze data to decompose search times into three behaviorally defined epochs: search initiation time, scanning time, and verification time (Malcolm & Henderson, 2009; Nuthmann, 2014; Spotorno et al., 2015). This was done to test how the availability of color in central as opposed to peripheral vision affects different subprocesses of search. For completeness, we note that the following eye-movement measures were highly correlated with search times, and therefore all showed the same qualitative pattern of

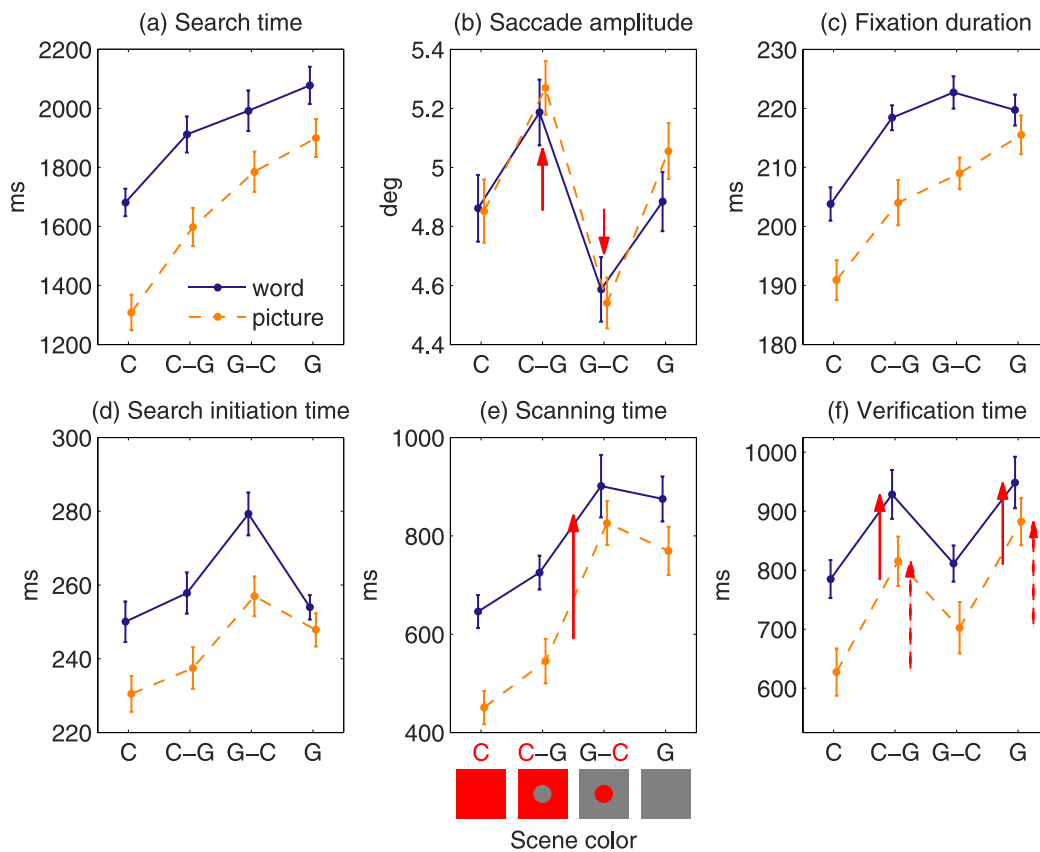


Figure 2. Performance and perception during visual search in real-world scenes. Each panel presents means obtained for a designated dependent variable (see panel title) as a function of scene color (x-axis). In addition, search guided by word cues (dark blue solid lines) is contrasted with search guided by picture cues (orange dashed lines). Error bars are within-subject standard errors, using the method described by Cousineau (2005). The red arrows highlight differences between conditions. Note that the sum of search initiation time (d), scanning time (e), and verification time (f) equals the search-time measure (a).

results: number of eye fixations needed to find the target, scan pattern ratio as a global measure of how directly the eyes moved to the search target (Brockmole & Henderson, 2006), and the proportion of the scene area sampled with central vision.

Gaze-based decomposition of search time

Search initiation time is the elapsed time between search scene onset and the first saccade (Malcolm & Henderson, 2009). It corresponds to initial saccade latency and measures the time needed to begin search. Previous research reported longer search initiation times for full-grayscale than for full-color scenes (Hwang et al., 2007). The literature reports no difference in initiation times as a function of cue specificity (Malcolm & Henderson, 2009, 2010) or a slight tendency for a picture cue advantage (Spotorno et al., 2014). In the present study, we used simple coding to test for effects of scene color (reference: full color) and cue specificity (reference: word cues) on search initiation times. The intercept in the color-cue model reflects the grand mean (i.e., the mean of the

eight cell means; $b = 251.7$ ms, $SE = 8.9$, $t = 28.34$). Independent of the color manipulation, search initiation was facilitated if search was guided by a picture cue rather than a word cue ($b = -17.4$ ms, $SE = 4.9$, $t = -3.55$). There was no significant difference in mean search initiation time between the full-color condition and the condition in which color was removed from central vision (C-G – C: $b = 7.5$ ms, $SE = 6.5$, $t = 1.15$). However, it took longer to initiate search when color was removed from peripheral vision (G-C – C: $b = 27.4$ ms, $SE = 6.6$, $t = 4.17$). Search initiation times did not differ significantly for color and grayscale scenes (G – C: $b = 10.4$ ms, $SE = 5.8$, $t = 1.78$). To facilitate comparison with the results by Hwang et al. (2007), who cued scene patches, we also conducted separate analyses for picture and word cues. The results qualitatively replicated the color-cue analysis with one exception: When search was guided by a picture cue, search initiation times were significantly longer for full-grayscale than for full-color scenes (G – C: $b = 16.4$, $SE = 7.6$, $t = 2.15$), which is in agreement with the results by Hwang et al. (2007).

Scanning time is the elapsed time between the first saccade (the end of the search initiation epoch) and the first fixation on the target object (Malcolm & Henderson, 2009). The scanning time measure reflects the actual search process. User-defined contrast coding was used to test our hypothesis that scanning times should be prolonged when color is not available in peripheral vision (G-C and G conditions). Accordingly, the first contrast compared the C and C-G conditions with the G-C and G conditions. Two additional contrasts were included as control variables: One compared the two conditions in which color was available in the periphery (C and C-G), and the other compared the two conditions in which color was removed from the periphery (G-C and G). In the color-cue LMM, the estimated grand mean was 727.1 ms ($SE = 64.6$, $t = 11.26$). Indeed, when color was not available in peripheral vision, scanning times were significantly prolonged (C and C-G vs. G-C and G: $b = 260.7$ ms, $SE = 41.2$, $t = 6.33$; see red arrow in Figure 2e). There was no significant difference between the two conditions in which color was removed from the periphery (G – G-C: $b = -35.3$ ms, $SE = 55.3$, $t = -0.64$). Scanning times were somewhat increased in the C-G condition compared with the C condition ($b = 99.9$ ms, $SE = 48.7$, $t = 2.05$). Finally, there was a significant speedup when picture cues rather than word cues were provided ($b = -130$ ms, $SE = 36.8$, $t = -3.54$).

Verification time is the elapsed time between the beginning of the first fixation on the target object and search termination (i.e., a button press). Verification time reflects the time needed to decide that the fixated object is the target. User-defined contrast coding was used to test our hypothesis that target verification is facilitated when color is available in central vision (C and G-C conditions). Therefore, the main contrast compared the C and G-C conditions with the C-G and G conditions. A second contrast compared the two conditions in which color was available in central vision (C and G-C), whereas the third contrast compared the two conditions in which color was removed from central vision (C-G and G). In the LMM, the estimated grand mean was 815.1 ms ($SE = 50.1$, $t = 16.26$). Indeed, when color was not available in central vision, verification times were significantly prolonged (C and G-C vs. C-G and G: $b = 166.6$ ms, $SE = 36.5$, $t = 4.56$; see red arrows in Figure 2f). There was no significant difference between the two conditions in which color was intact in central vision (G-C – C, $b = 58.1$ ms, $SE = 39.3$, $t = 1.48$). Likewise, there was no significant difference between the two conditions in which color was removed from central vision (G – C-G, $b = 42.9$ ms, $SE = 52.2$, $t = 0.82$). Finally, there was a significant speedup in target verification when the target was cued

with a picture rather than a verbal label ($b = -113.1$ ms, $SE = 41.5$, $t = -2.72$).

Saccade amplitudes and fixation durations

Saccade amplitudes and fixation durations were analyzed to further characterize eye-movement behavior during visual search. Saccade amplitudes scale with the size of the scene (von Wartburg et al., 2007), but they are generally considered to be an index of the spatial extent of parafoveal processing before a saccade is executed. Given the literature reviewed above, we do not expect saccade amplitudes in the full-color and full-grayscale conditions to differ. Studies that selectively filtered scene content in either peripheral or central vision consistently reported that observers show a tendency to fixate more locations in the untouched scene area and fewer in the degraded area (Laubrock, Cajar, & Engbert, 2013; Nuthmann, 2014). Therefore, smaller saccade amplitudes are expected when the availability of color is restricted to central vision (G-C), reflecting an increased tendency to place the eyes inside the moving window where color is preserved. Conversely, when the availability of color is restricted to peripheral vision (C-G), observers are expected to make somewhat larger saccades, reflecting an increased tendency to target the colored periphery. We do not expect an effect of cue specificity on saccade amplitudes.

The zigzag pattern presented in Figure 2b adheres to the predictions. For statistical evaluation we report results from a Color model, in which the full-color condition (C) served as the reference group (simple coding). The model included by-subject and by-item random intercepts; inclusion of random slopes did not improve the model fit. The fixed-effect intercept in the model reflects the grand mean for saccade amplitude (degrees; $b = 4.8$, $SE = 0.1$, $t = 45.82$). In the full-grayscale condition, saccade amplitudes did not differ from the full-color condition ($b = 0.1$, $SE = 0.1$, $t = 1.02$). However, when color was restricted to the periphery (C-G), saccade amplitudes were significantly increased ($b = 0.4$, $SE = 0.1$, $t = 3.5$). Conversely, when color was available only in central vision (G-C), saccade amplitudes were significantly reduced ($b = -0.3$, $SE = 0.1$, $t = -2.92$).

Fixation durations during scene exploration adapt to ongoing perceptual and cognitive processing (for a review, see Nuthmann, Smith, Engbert, & Henderson, 2010) and reflect difficulty in saccade planning (van Diepen & d'Ydewalle, 2003). Both foveal/central vision and peripheral vision play a critical role in regulating fixation duration during real-world scene perception and search (Laubrock et al., 2013). Previous research comparing full-color and full-grayscale scenes has reported conflicting evidence on mean fixation duration: no difference in a difficult scene-patch search task (Hwang et

al., 2007) as opposed to longer fixation durations for grayscale than for color scenes in scene memorization (von Wartburg et al., 2005) and free viewing (Ho-Phuoc et al., 2012). Fixation durations were expected to be shorter for picture-cued trials than for word-cued trials (Malcolm & Henderson, 2009, 2010). As with search times, the effect of scene color on fixation durations was tested using backward difference coding, and the effect of cue specificity was tested using simple coding. LMMs were conducted on log transformed fixation durations. Fixation durations were significantly shorter when search was guided by picture cues as opposed to word cues ($b = -0.049$, $SE = 0.013$, $t = -3.75$). With regard to the scene color manipulation, Figure 2c shows that fixation durations numerically increase across the four color conditions. When color was removed from central vision (C-G), fixation durations were significantly prolonged ($b = 0.053$, $SE = 0.014$, $t = 3.81$) compared with when the entire scene was presented in color (C). Removing color from the periphery (G-C) as opposed to central vision (C-G) was also associated with a significant increase in fixation duration ($b = 0.026$, $SE = 0.013$, $t = 1.99$). Fixation durations did not differ between conditions in which color was removed from the entire scene as opposed to the periphery only ($G - G-C$; $b = 0.006$, $SE = 0.012$, $t = 0.53$).

General discussion

The primary goal of the present study was to investigate the role color plays across the visual field when participants search for objects in scenes. To this end, we complemented conditions in which the entire scene stimulus was presented either in color or in grayscale with experimental conditions in which color information was selectively removed from either central or peripheral vision using the gaze-contingent moving window technique (Loschky & McConkie, 2002; McConkie & Rayner, 1975; Nuthmann, 2013). Specifically, we examined how color information influences different subbehaviors of search. In addition, we explored whether effects of scene color were modulated by the specificity of the target cue.

The influence of scene color and cue specificity was assessed by comparing classic eye-movement measures (fixation durations, saccade amplitudes), search times, and the subbehaviors that lead to the overall search time, including search initiation, scanning time, and verifications processes (Castelhano et al., 2008; Malcolm & Henderson, 2009; Nuthmann, 2014; Spotorno et al., 2015). All measures consistently showed that the availability of either color information or a specific target template facilitated search and that the effects of color and cue did not interact.

Search times steadily increased (numerically) as the amount of color information in the scene decreased (Figure 2a). Relative to the full-color condition, search times were significantly prolonged in the C-G, G-C, and full-grayscale conditions, respectively. Among the latter three conditions, search-time differences were somewhat less pronounced. A gaze-data based segmentation of search time revealed color-mediated effects on specific subprocesses of search. First, effects of scene color on search initiation times (Figure 2d) were largely limited to increased latencies in the condition in which color was removed from peripheral vision (G-C). We speculate that this may represent a tendency of the eyes to stay longer at the initial location where color is available. Moreover, search initiation times were significantly longer for full-grayscale scenes than for full-color scenes, but only for pictures cues (cf. Hwang et al., 2007) and not for word cues. Second, the data lent support to our main hypothesis that the availability of color in peripheral vision may facilitate target localization, whereas the availability of color in central vision may facilitate target verification. On one hand, when color was not available in peripheral vision (G-C and G), scanning time was prolonged (Figure 2e), which means that it took longer to locate the target in the scene. On the other hand, when color was not available in central vision (C-G and G), the process of verifying the identity of the target took longer to complete (Figure 2f). Notably, the differential effects of scene color on search initiation time, scanning time, and verification time combine to produce a steady increase in search times across the four color conditions. It is important to note that color effects for the various subprocesses of search operate on different scales. The fixed-effect regression coefficients, quantifying differences between conditions, showed that temporal differences were larger for scanning time (261 ms) than for verification time (167 ms) and much smaller for search initiation time (27 ms).

We first compare the full-color condition with the C-G condition. When color was removed from central vision (C-G), search times were increased. This effect is driven mainly by increased verification times but also by some increase in scanning times. Next, we compare the full-grayscale condition with the G-C condition, in which color remained intact in central vision. Increased verification times for full-grayscale scenes boost the search times in this condition over the G-C condition. This is counteracted by increased initiation times in the G-C condition. As a result, the numerical search-time difference between the two conditions ($G - G-C$: 112 ms) was not significant. We then compare the results from the two moving window conditions. Restricting color to peripheral vision (C-G) impedes the verification process but aids the localization process, though target localization is not quite as fast as in the full-color

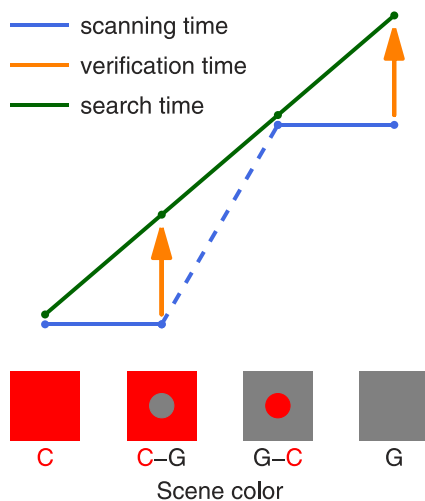


Figure 3. Schematic visualization of how differential effects of scene color on scanning times and verification times combine to produce a steady increase in search times across the four color conditions. The availability of color in peripheral vision (C and C-G conditions) facilitates target localization, whereas the availability of color in central vision (C and G-C conditions) facilitates target verification. Therefore, when color is not available in peripheral vision (G-C and G conditions), scanning times are prolonged (blue “step function”). In contrast, verification times are prolonged when color is not available in central vision (C-G and G conditions; orange arrows).

condition. Conversely, restricting color to central vision (G-C) slows down search initiation and the localization process but facilitates the verification process. Those opposing effects combine to produce longer search times in the G-C than in the C-G condition (G-C – C-G: 131 ms). Finally, the clear disadvantage of full-grayscale scenes over full-color scenes originates from increases in both scanning time and verification time. Figure 3 provides a simplified schematic visualization of these interrelationships.

Next, we compare our results with results from a previous study, which investigated the importance of foveal, central, and peripheral vision for object-in-scene search (Nuthmann, 2014). In this study, a given region of the visual field was degraded by applying a severe, above-threshold low-pass filter that removed fine detail information about objects in the scene while maintaining global luminance and color information (Nuthmann, 2014). One condition (Large Blindspot) simulated the absence of central vision, and another one (Large Spotlight) simulated the absence of peripheral vision. In both conditions, search times were elevated compared with a natural vision baseline condition. Notably, the search-time costs originated from different subprocesses of search. When peripheral vision was withheld it took longer to localize the search object in space. In contrast, when central vision was denied it took longer to verify the identity of the target.

A similar pattern of results was found in the present study, in which we manipulated one specific image feature—namely, color. Removing color from peripheral vision slowed down target localization. In contrast, removing color from central vision slowed down target verification. Target localization was not quite as fast as in the full-color condition but faster than in conditions in which color was removed from peripheral vision. Therefore, the data suggest that color in the center and in the periphery does not serve completely different subprocesses of search.

Moreover, our finding that color in central vision facilitates the recognition of the target is in agreement with surface-plus-edge theories of object recognition, which suggest that both edges and color contribute to recognition (for a review, see Bramao, Reis, Petersson, & Faisca, 2011). At the same time, the present data do not support traditional theories of object recognition that emphasize the importance of shape and de-emphasize the role of color (Biederman, 1987).

Removing color information from the scene increased fixation durations (Figure 2c). The result that fixation durations were longer for full-grayscale than for full-color scenes is in agreement with findings from other scene-viewing tasks (Ho-Phuoc et al., 2012; von Wartburg et al., 2005) but contrasts with the null effect observed by Hwang et al. (2007) in a scene search task. Selectively removing color from either central vision (C-G) or peripheral vision (G-C) led to increased fixation durations, supporting the view that color information in both central and peripheral vision plays a critical role in regulating fixation durations. Our findings add to the literature suggesting that fixation durations are sensitive to lower level stimulus properties such as the luminance of the scene (Henderson, Nuthmann, & Luke, 2013; Loftus, 1985; Walshe & Nuthmann, 2014).

Manipulating the availability of color information in either central or peripheral vision led to systematic adjustments of mean saccade amplitudes. Compared to the full-color (C) and full-grayscale (G) conditions, saccade amplitudes were reduced when color information was restricted to central vision (G-C) and increased when color was available only in peripheral vision (C-G); see red arrows in Figure 2b. This data pattern reflects a tendency of sending the eyes to scene areas where color was available (e.g., in the C-G condition, the eyes were drawn toward the colored periphery), which is in agreement with the general finding that observers have a tendency to fixate more locations in the nondegraded scene area than the degraded area (Laubrock et al., 2013; Nuthmann, 2014).

As a secondary question, we investigated the relative contribution of color information across the visual field to target template search. Specific target templates were found to facilitate search but did not interact with the presence of color in the search

scenes. Search times were shorter when the target was cued with a picture than when it was cued with a verbal label (Figure 2a), replicating previous research (Malcolm & Henderson, 2009, 2010; Spotorno et al., 2014). The speedup for picture cues was observed for all three subprocesses of search (Figure 2d through f). The finding that a more specific target template facilitates scanning and verification is consistent with the literature. Regarding search initiation times, previous experiments found no difference between picture and word cues (Malcolm & Henderson, 2009, 2010). In the present study, however, search initiation was facilitated by a picture cue (see also Spotorno et al., 2014, reporting a marginally significant effect), thereby supporting the idea that search initiation time can be a sensitive measure when studying target template guidance. Furthermore, picture cue trials were associated with shorter fixation durations than word cue trials (Figure 2c), suggesting that a more specific target template facilitates faster scene processing (Malcolm & Henderson, 2009, 2010).

The results suggest that viewers rely on color as well as other features of a target template. While picture cues always benefited search more than word cues, and full-color scenes always benefited search more than full-grayscale scenes, these two manipulations never interacted but rather formed an additive effect. This suggests that color is generally needed for the maximum facilitation of target template guided search, but in its absence other features (e.g., edges, luminance) can be used. Fully exploring the role of target template color and its relation with scene color is an interesting avenue for future study (see Malcolm & Henderson, 2011).

Finally, we chose to use LMMs over analysis of variance (ANOVA) to analyze our data because its many advantages in experimental research are well documented (Kliegl et al., 2011). While there is a strong movement toward replacing ANOVA with LMMs in psycholinguistics (Baayen et al., 2008; Cunnings, 2012; Locker, Hoffman, & Bovaird, 2007), researchers in real-world scene perception and search have only just begun to exploit LMMs (e.g., Nuthmann & Einhäuser, 2015; Spotorno et al., 2015). One advantage of using LMMs in the present context is that the fixed-effect coefficients are directly interpretable because they describe differences between specific conditions or clusters of conditions.

Conclusions

The present results extend the existing literature that has highlighted the role of color in scene segmentation

and object recognition (Gegenfurtner, 2003). As a novel contribution, we used eye tracking to investigate the importance of color in central and peripheral vision for particular subprocesses of object-in-scene search. The main finding was that the availability of color in peripheral vision facilitates the process of localizing the target in space, whereas the availability of color in central vision facilitates the process of verifying the identity of the target.

Keywords: naturalistic scenes, visual search, color, eye movements, peripheral vision

Acknowledgments

Portions of this research were presented at the 17th European Conference on Eye Movements in Lund, Sweden, August 2013. The authors thank Thomas Dixon for assistance with data collection.

Commercial relationships: none.

Corresponding author: Antje Nuthmann.

Email: Antje.Nuthmann@ed.ac.uk.

Address: Psychology Department, School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Edinburgh, United Kingdom.

Footnote

¹ We note that this effect was significant when search times were log transformed ($b = 1.09$, $SE = 1.03$, $t = 2.78$).

References

- Amano, K., & Foster, D. H. (2014). Influence of local scene color on fixation position in visual search. *Journal of the Optical Society of America A: Optics, Image Science, and Vision*, 31(4), A254–A262, doi: 10.1364/josaa.31.00a254.
- Amano, K., Foster, D. H., Mould, M. S., & Oakley, J. P. (2012). Visual search in natural scenes explained by local color properties. *Journal of the Optical Society of America A: Optics, Image Science, and Vision*, 29(2), A194–A199, doi:10.1364/JOSAA.29.00A194.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and*

- Language*, 59(4), 390–412, doi:10.1016/j.jml.2007.12.005.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278, doi:10.1016/j.jml.2012.11.001.
- Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2015). *lme4: Linear mixed-effects models using Eigen and S4* (R package version 1.1-8). Retrieved from <http://CRAN.R-project.org/package=lme4>
- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94(2), 115–147, doi:10.1037/0033-295x.94.2.115.
- Biederman, I., & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception and Performance*, 19(6), 1162–1182, doi:10.1037/0096-1523.19.6.1162.
- Biederman, I., & Ju, G. (1988). Surface versus edge-based determinants of visual recognition. *Cognitive Psychology*, 20(1), 38–64, doi:10.1016/0010-0285(88)90024-2.
- Borji, A., & Itti, L. (2013). State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1), 185–207, doi:10.1109/tpami.2012.89.
- Borji, A., Sihite, D. N., & Itti, L. (2013). Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study. *IEEE Transactions on Image Processing*, 22(1), 55–69, doi:10.1109/tip.2012.2210727.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433–436, doi:10.1163/156856897X00357.
- Bramao, I., Reis, A., Petersson, K. M., & Faisca, L. (2011). The role of color information on object recognition: A review and meta-analysis. *Acta Psychologica*, 138(1), 244–253, doi:10.1016/j.actpsy.2011.06.010.
- Bravo, M. J., & Farid, H. (2009). The specificity of the search template. *Journal of Vision*, 9(1):34, 1–9, doi:10.1167/9.1.34. [PubMed] [Article]
- Brockmole, J. R., & Henderson, J. M. (2006). Recognition and attention guidance during contextual cueing in real-world scenes: Evidence from eye movements. *The Quarterly Journal of Experimental Psychology*, 59(7), 1177–1187, doi:10.1080/17470210600665996.
- Castelhano, M. S., Pollatsek, A., & Cave, K. R. (2008). Typicality aids search for an unspecified target, but only in identification and not in attentional guidance. *Psychonomic Bulletin & Review*, 15(4), 795–801, doi:10.3758/pbr.15.4.795.
- Cornelissen, F. W., Bruin, K. J., & Kooijman, A. C. (2005). The influence of artificial scotomas on eye movements during visual search. *Optometry and Vision Science*, 82(1), 27–35.
- Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, & Computers*, 34(4), 613–617, doi:10.3758/BF03195489.
- Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorial in Quantitative Methods for Psychology*, 1(1), 42–45.
- Cummings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, 28(3), 369–382, doi:10.1177/0267658312443651.
- D'Zmura, M. (1991). Color in visual search. *Vision Research*, 31(6), 951–966, doi:10.1016/0042-6989(91)90203-h.
- Davidoff, J. B., & Ostergaard, A. L. (1988). The role of colour in categorial judgements. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 40(3), 533–544, doi:10.1080/02724988843000069.
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18, 193–222, doi:10.1146/annurev.neuro.18.1.193.
- Eckstein, M. P., Drescher, B. A., & Shimozaki, S. S. (2006). Attentional cues in real scenes, saccadic targeting, and Bayesian priors. *Psychological Science*, 17(11), 973–980, doi:10.1111/j.1467-9280.2006.01815.x.
- Foulsham, T., & Kingstone, A. (2013). Optimal and preferred eye landing positions in objects and scenes. *Quarterly Journal of Experimental Psychology*, 66(9), 1707–1728, doi:10.1080/17470218.2012.762798.
- Foulsham, T., & Underwood, G. (2011). If visual saliency predicts search, then why? Evidence from normal and gaze-contingent search tasks in natural scenes. *Cognitive Computation*, 3(1), 48–63, doi:10.1007/s12559-010-9069-9.
- Frey, H. P., Honey, C., & König, P. (2008). What's color got to do with it? The influence of color on visual attention in different categories. *Journal of Vision*, 8(14):6, 1–17, doi:10.1167/8.14.6. [PubMed] [Article]

- Frey, H. P., Wirz, K., Willenbockel, V., Betz, T., Schreiber, C., Troscianko, T., & König, P. (2011). Beyond correlation: Do color features influence attention in rainforest? *Frontiers in Human Neuroscience*, 5, 36, doi:10.3389/fnhum.2011.00036.
- Gegenfurtner, K. R. (2003). Cortical mechanisms of colour vision. *Nature Reviews Neuroscience*, 4(7), 563–572, doi:10.1038/nrn1138.
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York, NY: Cambridge University Press.
- Hannus, A., van den Berg, R., Bekkering, H., Roerdink, J. B. T. M., & Cornelissen, F. W. (2006). Visual search near threshold: Some features are more equal than others. *Journal of Vision*, 6(4), 523–540, doi:10.1167/6.4.15. [PubMed] [Article]
- Hansen, T., Pracejus, L., & Gegenfurtner, K. R. (2009). Color perception in the intermediate periphery of the visual field. *Journal of Vision*, 9(4):26, 1–12, doi:10.1167/9.4.26. [PubMed] [Article]
- Henderson, J. M., Nuthmann, A., & Luke, S. G. (2013). Eye movement control during scene viewing: Immediate effects of scene luminance on fixation durations. *Journal of Experimental Psychology: Human Perception and Performance*, 39(2), 318–322, doi:10.1037/a0031224.
- Ho-Phuoc, T., Guyader, N., Landragin, F., & Guerin-Dugue, A. (2012). When viewing natural scenes, do abnormal colors impact on spatial or temporal parameters of eye movements? *Journal of Vision*, 12(2):4, 1–13, doi:10.1167/12.2.4. [PubMed] [Article]
- Humphrey, G. K., Goodale, M. A., Jakobson, L. S., & Servos, P. (1994). The role of surface information in object recognition: Studies of a visual form agnostic and normal subjects. *Perception*, 23(12), 1457–1481, doi:10.1068/p231457.
- Hwang, A. D., Higgins, E. C., & Pomplun, M. (2007). How chromaticity guides visual search in real-world scenes. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Cognitive Science Society* (pp. 371–376). Austin, TX: Cognitive Science Society.
- Hwang, A. D., Higgins, E. C., & Pomplun, M. (2009). A model of top-down attentional control during visual search in complex scenes. *Journal of Vision*, 9(5):25, 1–18, doi:10.1167/9.5.25. [PubMed] [Article]
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40(10–12), 1489–1506, doi:10.1016/S0042-6989(99)00163-7.
- Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11), 1254–1259, doi:10.1109/34.730558.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59(4), 434–446, doi:10.1016/j.jml.2007.11.007.
- Jost, T., Ouerhani, N., von Wartburg, R., Müri, R., & Hügli, H. (2005). Assessing the contribution of color in visual attention. *Computer Vision and Image Understanding*, 100(1–2), 107–123, doi:10.1016/j.cviu.2004.10.009.
- Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception*, 36, 14.
- Kliegl, R., Wei, P., Dambacher, M., Yan, M., & Zhou, X. (2011). Experimental effects and individual differences in linear mixed models: Estimating the relationship between spatial, object, and attraction effects in visual attention. *Frontiers in Psychology*, 1, 1–12, doi:10.3389/fpsyg.2010.00238.
- Laubrock, J., Cajar, A., & Engbert, R. (2013). Control of fixation duration during scene viewing by interaction of foveal and peripheral processing. *Journal of Vision*, 13(12):11, 1–20, doi:10.1167/13.12.11. [PubMed] [Article]
- Locker, L., Hoffman, L., & Bovaird, J. A. (2007). On the use of multilevel modeling as an alternative to items analysis in psycholinguistic research. *Behavior Research Methods*, 39(4), 723–730, doi:10.3758/bf03192962.
- Loftus, G. R. (1985). Picture perception: Effects of luminance on available information and information-extraction rate. *Journal of Experimental Psychology: General*, 114(3), 342–356, doi:10.1037/0096-3445.114.3.342.
- Loschky, L. C., & McConkie, G. W. (2002). Investigating spatial vision and dynamic attentional selection using a gaze-contingent multiresolutional display. *Journal of Experimental Psychology: Applied*, 8(2), 99–117, doi:10.1037/1076-898X.8.2.99.
- Malcolm, G. L., & Henderson, J. M. (2009). The effects of target template specificity on visual search in real-world scenes: Evidence from eye movements. *Journal of Vision*, 9(11):8, 1–13, doi:10.1167/9.11.8. [PubMed] [Article]
- Malcolm, G. L., & Henderson, J. M. (2010). Combining top-down processes to guide eye movements during real-world scene search. *Journal of Vision*, 10(2):4, 1–11, doi:10.1167/10.2.4. [PubMed] [Article]
- Malcolm, G. L., & Henderson, J. M. (2011). Visual

- search in real-world scenes: The role of color during target template guidance. *Journal of Eye Movement Research*, 4(3), 64.
- Malcolm, G. L., & Shomstein, S. (2015). Object-based attention in real-world scenes. *Journal of Experimental Psychology: General*, 144(2), 257–263, doi:10.1037/xge0000060.
- McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics*, 17(6), 578–586, doi:10.3758/BF03203972.
- Mullen, K. T. (1991). Colour vision as a post-receptoral specialization of the central visual field. *Vision Research*, 31(1), 119–130, doi:10.1016/0042-6989(91)90079-k.
- Mullen, K. T., & Kingdom, F. A. A. (1996). Losses in peripheral colour sensitivity predicted from “hit and miss” post-receptoral cone connections. *Vision Research*, 36(13), 1995–2000, doi:10.1016/0042-6989(95)00261-8.
- Nagy, A. L., & Sanchez, R. R. (1990). Critical color differences determined with a visual search task. *Journal of the Optical Society of America A: Optics, Image Science, and Vision*, 7(7), 1209–1217, doi:10.1364/josaa.7.001209.
- Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research*, 46(5), 614–621, doi:10.1016/j.visres.2005.08.025.
- Nuthmann, A. (2013). On the visual span during object search in real-world scenes. *Visual Cognition*, 21(7), 803–837, doi:10.1080/13506285.2013.832449.
- Nuthmann, A. (2014). How do the regions of the visual field contribute to object search in real-world scenes? Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, 40(1), 342–360, doi:10.1037/a0033854.
- Nuthmann, A., & Einhäuser, W. (2015). A new approach to modeling the influence of image features on fixation selection in scenes. *Annals of the New York Academy of Sciences*, 1339, 82–96, doi:10.1111/nyas.12705.
- Nuthmann, A., & Henderson, J. M. (2010). Object-based attentional selection in scene viewing. *Journal of Vision*, 10(8):20, 1–19, doi:10.1167/10.8.20. [PubMed] [Article]
- Nuthmann, A., Smith, T. J., Engbert, R., & Henderson, J. M. (2010). CRISP: A computational model of fixation durations in scene viewing. *Psychological Review*, 117(2), 382–405, doi:10.1037/a0018924.
- Ostergaard, A. L., & Davidoff, J. B. (1985). Some effects of color on naming and recognition of objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11(3), 579–587, doi:10.1037/0278-7393.11.3.579.
- Pajak, M., & Nuthmann, A. (2013). Object-based saccadic selection during scene perception: Evidence from viewing position effects. *Journal of Vision*, 13(5):2, 1–21, doi:10.1167/13.5.2. [PubMed] [Article]
- R Core Team. (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at <https://www.R-project.org/>.
- Rosenholtz, R., Li, Y. Z., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision*, 7(2):17, 1–22, doi:10.1167/7.2.17. [PubMed] [Article]
- Rutishauser, U., & Koch, C. (2007). Probabilistic modeling of eye movement data during conjunction search via feature-based attention. *Journal of Vision*, 7(6):5, 1–20, doi:10.1167/7.6.5. [PubMed] [Article]
- Spotorno, S., Malcolm, G. L., & Tatler, B. W. (2014). How context information and target information guide the eyes from the first epoch of search in real-world scenes. *Journal of Vision*, 14(2):7, 1–21, doi:10.1167/14.2.7. [PubMed] [Article]
- Spotorno, S., Malcolm, G. L., & Tatler, B. W. (2015). Disentangling the effects of spatial inconsistency of targets and distractors when searching in realistic scenes. *Journal of Vision*, 15(2):12, 1–21, doi:10.1167/15.2.12. [PubMed] [Article]
- Strasburger, H., Rentschler, I., & Jüttner, M. (2011). Peripheral vision and pattern recognition: A review. *Journal of Vision*, 11(5):13, 1–82, doi:10.1167/11.5.13. [PubMed] [Article]
- Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: Effects of scale and time. *Vision Research*, 45(5), 643–659, doi:10.1016/j.visres.2004.09.017.
- Theeuwes, J. (1994). Stimulus-driven capture and attentional set: Selective search for color and visual abrupt onsets. *Journal of Experimental Psychology: Human Perception and Performance*, 20(4), 799–806, doi:10.1037//0096-1523.20.4.799.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1), 97–136, doi:10.1016/0010-0285(80)90005-5.
- UCLA Statistical Consulting Group. (2011). *R Library: Contrast coding systems for categorical variables*. Retrieved from http://www.ats.ucla.edu/stat/r/library/contrast_coding.htm

- van Diepen, P. M. J., & d'Ydewalle, G. (2003). Early peripheral and foveal processing in fixations during scene perception. *Visual Cognition*, 10(1), 79–100, doi:10.1080/13506280143000023.
- von Wartburg, R., Ouerhani, N., Pflugshaupt, T., Nyffeler, T., Wurtz, P., Hügli, H., & Müri, R. M. (2005). The influence of colour on oculomotor behaviour during image perception. *Neuroreport*, 16(14), 1557–1560, doi:10.1097/01.wnr.0000180146.84020.c4.
- von Wartburg, R., Wurtz, P., Pflugshaupt, T., Nyffeler, T., Lüthi, M., & Müri, R. M. (2007). Size matters: Saccades during scene perception. *Perception*, 36(3), 355–365, doi:10.1068/p5552.
- Walshe, R. C., & Nuthmann, A. (2014). Asymmetrical control of fixation durations in scene viewing. *Vision Research*, 100, 38–46, doi:10.1016/j.visres.2014.03.012.
- Williams, L. G. (1967). The effects of target specification on objects fixated during visual search. *Acta Psychologica*, 27, 355–360, doi:10.1016/0001-6918(67)90080-7.
- Wilson, H. R., Levi, D., Maffei, L., Rovamo, J., & DeValois, R. (1990). The perception of form: Retina to striate cortex. In L. Spillmann & J. S. Werner (Eds.), *Visual perception: The neurophysiological foundations* (pp. 231–272). San Diego, CA: Academic Press.
- Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, 1(2), 202–238, doi:10.3758/bf03200774.
- Wurm, L. H., Legge, G. E., Isenberg, L. M., & Luebker, A. (1993). Color improves object recognition in normal and low vision. *Journal of Experimental Psychology: Human Perception and Performance*, 19(4), 899–911, doi:10.1037/0096-1523.19.4.899.
- Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review*, 115(4), 787–835, doi:10.1037/a0013118.